

## How to Import Data into Microsoft Access

This tutorial demonstrates how to import an Excel file into an Access database. You can also follow these same steps to import other data tables into Access, such as dbf files or txt files, with some minor variation. This document uses some standard database terminology, referring to columns as fields and rows as records.

Working with data in Microsoft Access databases can be preferable to working with spreadsheets for a number of reasons. One database file can hold several tables, which makes consolidating and organizing data easier. Databases are also great for calculating summaries and aggregating data more efficiently than a spreadsheet. Access is a relational database, which means that you can bring related data stored in different tables together to summarize and calculate data; a feat that a spreadsheet like Excel cannot perform. Databases are a necessity when working with large datasets, as spreadsheets have a limit of 65,000 records. Finally, Microsoft Access databases are compatible with ArcGIS, as tables within Access databases can be directly accessed by ArcGIS and added to ArcMap. Tables can then be joined to boundary files based on a common field, and the data within the tables can be mapped. This is highly preferable to working with table in the dbf file format, which has very particular syntax rules that makes them difficult to work with. Access databases can also serve as geodatabases for holding boundary files and data tables (for instructions on importing shapefiles into Access, see the *Import Shapefiles to Access* tutorial).

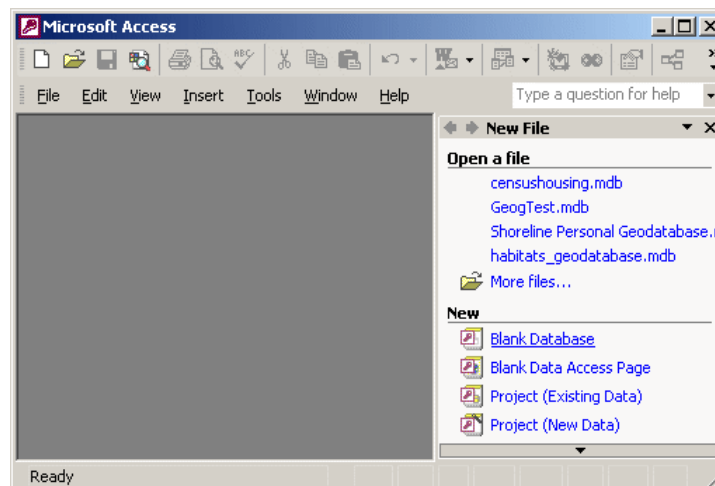
NOTE – if you plan to eventually use a database as a geodatabase, for holding geospatial data and tabular data, and you are using ArcGIS 9.2, you MUST create the database within Arc Catalog by right-clicking in the table of contents, clicking *new*, and *new personal geodatabase*. You cannot import geospatial data into an Access database that was not created with ArcGIS 9.2. If you plan on using the database for just holding and working with tabular data, or you are using an earlier version of ArcGIS (8.3 to 9.1), you can safely create a database within Access, Windows Explorer, or ArcGIS, and it will work as a geodatabase.

### Creating a New, Empty Database

The first step is to create a new, empty database. Unlike a spreadsheet, where you can launch the program and immediately start entering data, in Access you must create an empty database first before you can enter or import data. You can think of this empty database as a container that is going to hold all of your information. We also are unable to simply open an Excel file or any other file directly in Access. We have to import the file into a new or existing database.

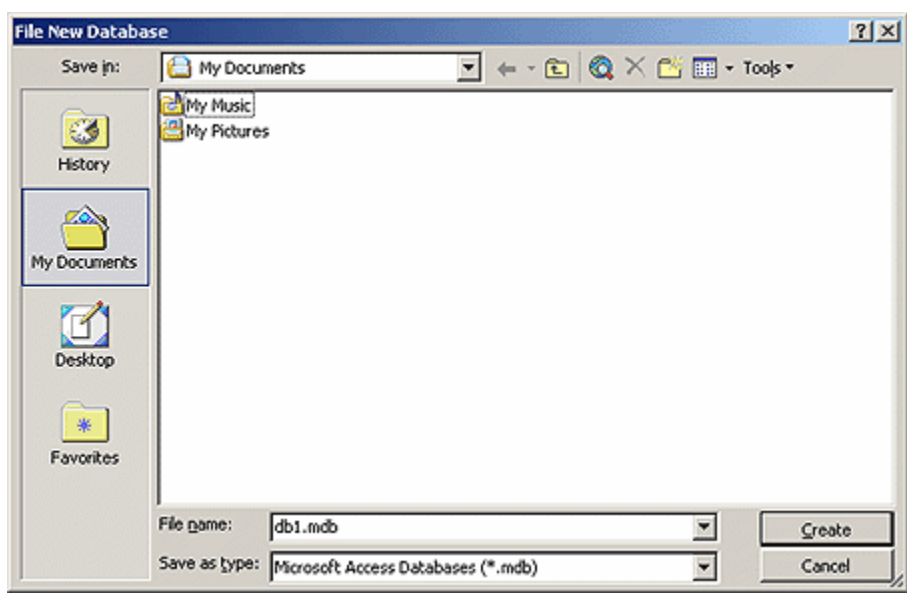
Once you open Access, depending on which version you are using, either click on *New Blank Database* in the menu on the right, or go to the *File* menu and select *New Database* (see Figure 1).

Figure 1



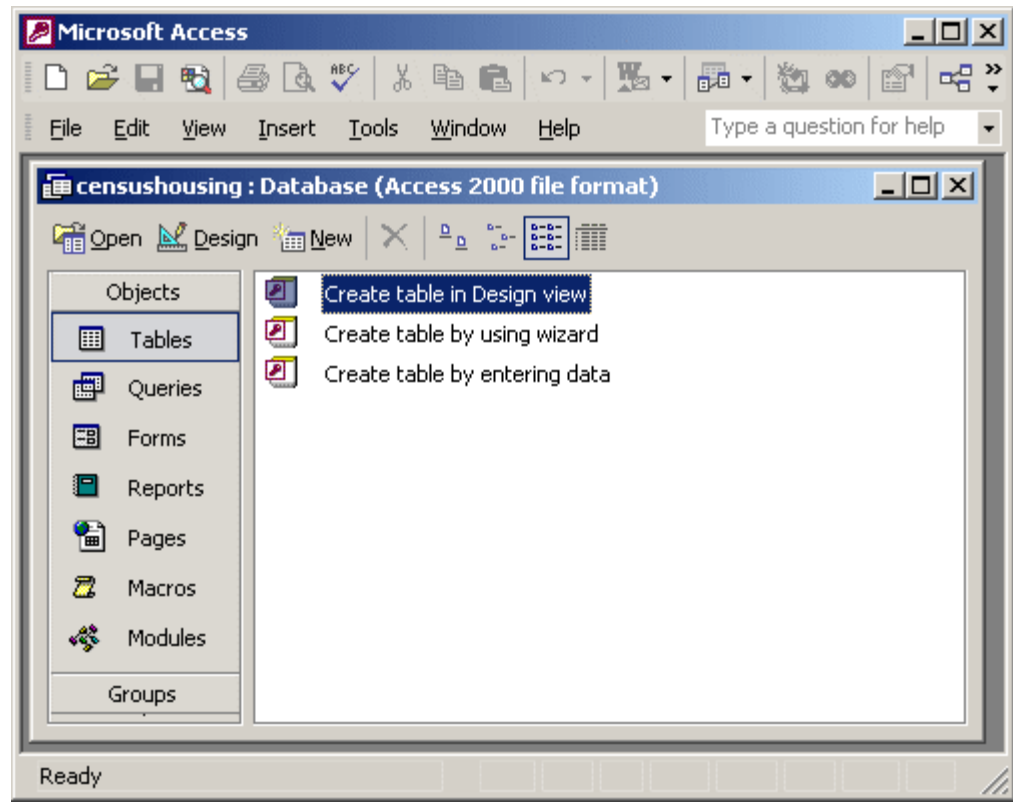
You will be prompted to name your new database file and navigate to a folder for storing it (Figure 2). Microsoft Access databases have the file extension mdb.

Figure 2



Once you create the database you will be presented with the database screen (Figure 3). By default you are in the *Tables* section (see the menu on the left), which is naturally blank, as you have not added any data yet. The top of the database window shows the name of the database. We have named it *cenushousing*, as we will be importing housing data from the US Census in this example.

Figure 3



Next to the name of the database is the Access file format version. By default, the version of Access that you are using defaults to the file format used by the previous version, to help insure backward compatibility between your datasets. In this example, we are using Access 2002, so by default it is storing the database in an Access 2000 format. Access 2000, 2002, and 2003 formats are cross-compatible with each other. If you are using Access 2007, you will have to convert these formats into the newer 2007 format.

### Preparing Data for Import

Before we begin importing data into the database, it is necessary to prepare our dataset for import. In this case, we have a Microsoft Excel file with housing data from the US Census, which was downloaded from the *American Factfinder* (see Figure 4).

Figure 4

	A	B	C	D	E	F
1	GEO_ID	GEO_ID2	SUMLEVEL	GEO_NAME	H003001	H003002
2	Geography Identifier	Geography Identifier	Geographic Summary Level	Geography	Housing units: Total	Housing units: Occupied
3	14000US53001950100	53001950100	140	Adams County,	1149	1029
4	0	53001950200	140	Adams County,	899	716
5	0	53001950300	140	Adams County,	1786	1628
6	0	53001950400	140	Adams County,	841	814
7	0	53001950500	140	Adams County,	1098	1042
8	0	53003980100	140	Asotin County,	1609	1419
9	0	53003980200	140	Asotin County,	1870	1765
10	0	53003980300	140	Asotin County,	1821	1640
11	0	53003980400	140	Asotin County,	888	809
12	0	53003980500	140	Asotin County,	1542	1437

In this example, we do have some preparatory work to do. In a database, and unlike a spreadsheet, all data in a particular column must be of the same type; numbers, text, dates, etc. If there is a column of numbers, such as housing units, then this column must contain numbers and no text whatsoever. Numbers can be stored as text in a database, but you will not be able to perform any calculations on these columns. Generally speaking, identification numbers should be saved as text, and numbers that you would perform mathematical calculations on should be stored as numbers. So the first step is to examine each of columns to verify that they contain the same data type, and make edits where appropriate.

Columns in a database can only have one field heading or name. The first row is recognized as the header row, and all subsequent rows are recognized as data. In this Excel table, we have two header rows. We have to delete one of them, otherwise we will get errors when we try to import the table into Access. In this case, it is best to keep the first row, which contains the abbreviated names and codes: GEO\_ID, GEO\_ID2, SUMLEVEL etc, and NOT the row that contains the full name: Geography Identifier, Geographic Summary Level, etc. Why? The full names are too long and may exceed the database's length limit for field names. They also contain spaces, which is something that is not allowed in field names. Field names

should never contain strange punctuation, and should never begin with numbers. Once we delete this row, our spreadsheet looks like Figure 5:

Figure 5

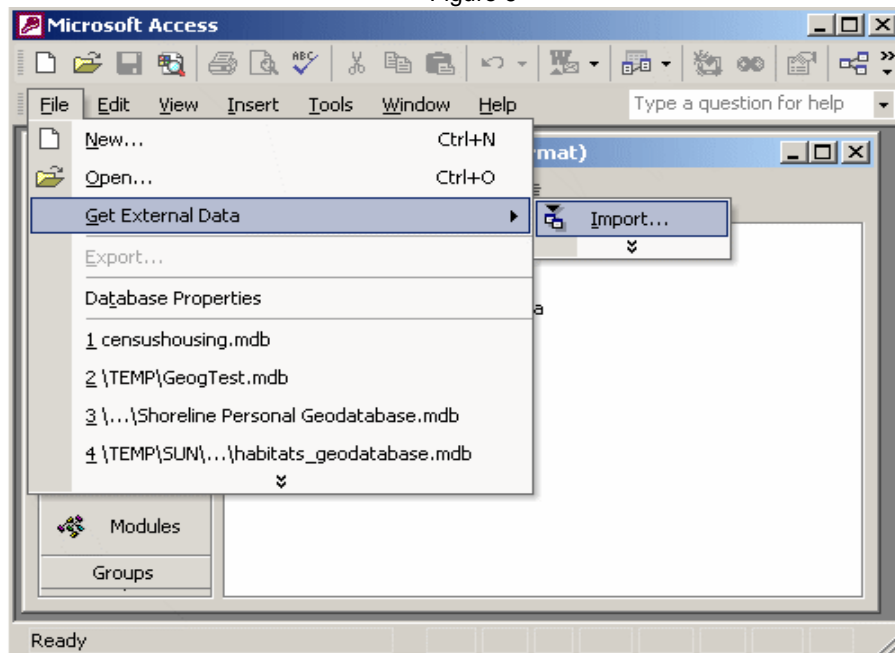
	A	B	C	D	E	F
	GEO_ID	GEO_ID2	SUMLEVEL	GEO_NAME	H003001	H003002
1						
2	14000US53001950100	53001950100	140	Adams County,	1149	1029
3	0	53001950200	140	Adams County,	899	716
4	0	53001950300	140	Adams County,	1786	1628
5	0	53001950400	140	Adams County,	841	814
6	0	53001950500	140	Adams County,	1098	1042
7	0	53003980100	140	Asotin County,	1609	1419
8	0	53003980200	140	Asotin County,	1870	1765
9	0	53003980300	140	Asotin County,	1821	1640
10	0	53003980400	140	Asotin County,	888	809
11	0	53003980500	140	Asotin County,	1542	1437
12	0	53003980600	140	Asotin County,	1381	1294
13	14000US53005010100	53005010100	140	Benton County,	1894	1839
14	14000US53005010201	53005010201	140	Benton County,	1430	1376

Save the changes, close the spreadsheet, and go back to the Access database. We are now ready to import our table.

### Importing Data

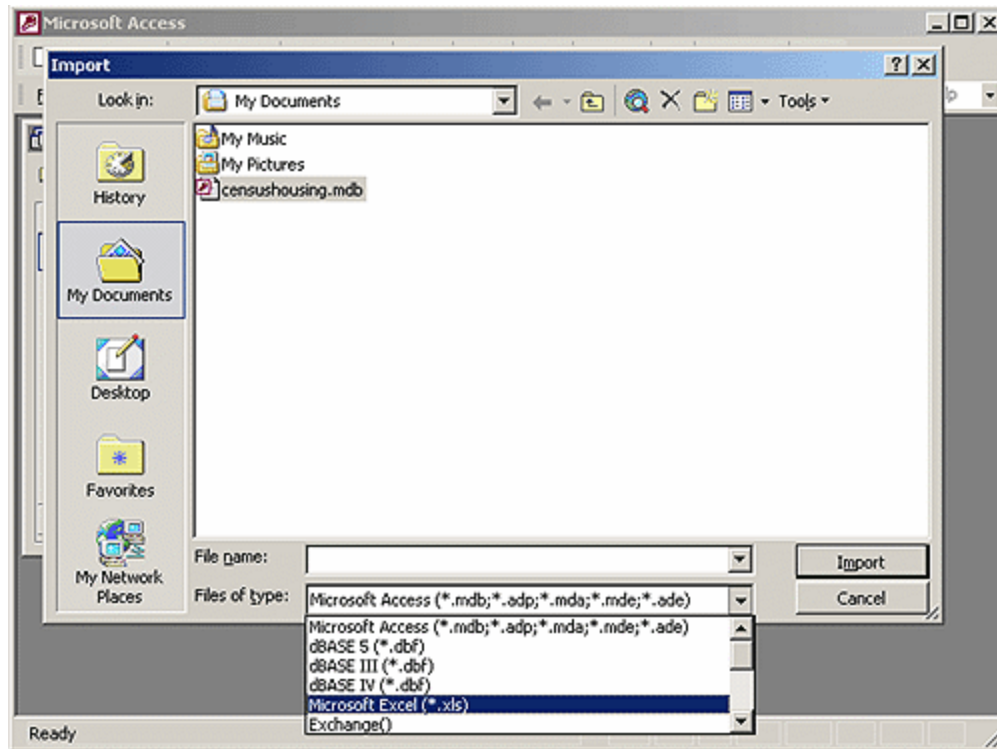
Go to the *File* menu, and select *Get External Data*, then *Import* (Figure 6).

Figure 6



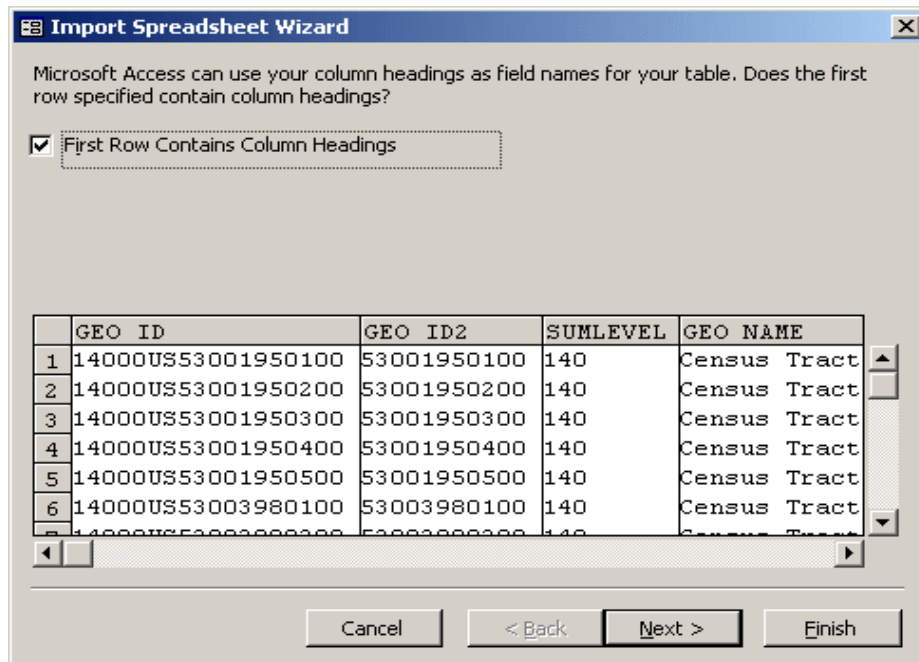
Next, we are prompted to select which file we want to import (Figure 7). By default, Access is looking for Access mdb files first. Change the file type to Excel files, browse to the folder where the file is stored, and select it.

Figure 7



Now the import process begins. If our Excel file contained multiple sheets, we would be prompted to specify which sheet to import. In this example, our Excel file only contains one sheet, so Access assumes that this is what we want to import. The first question is whether the first column in our spreadsheet contains the column headings (Figure 8). This is the case, so we check the box (if this was not the case, then Access would assign generic column headings for us).

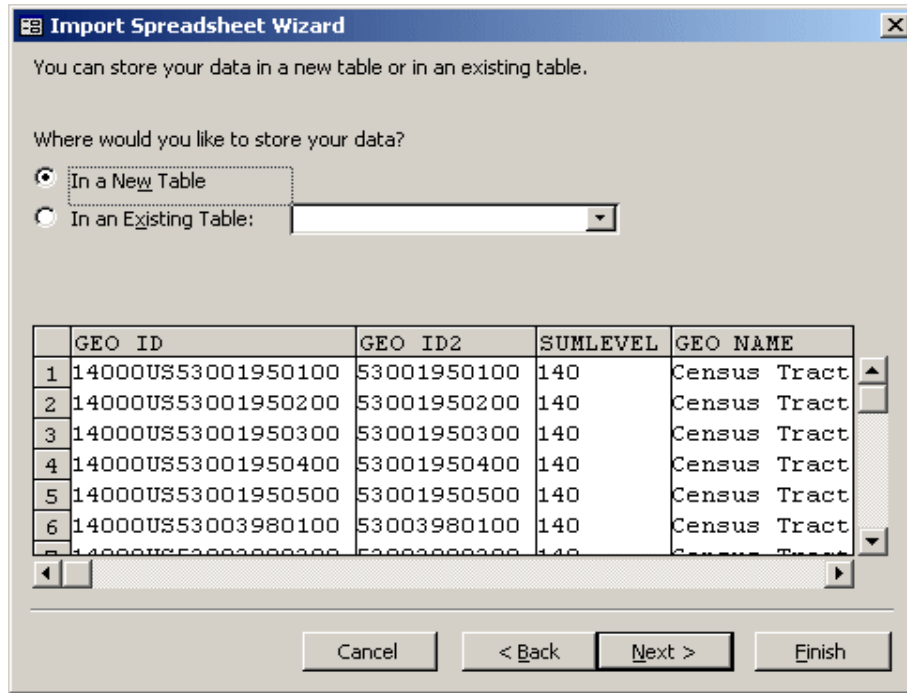
Figure 8



(NOTE – If we were trying to import another type of file, such as a text file, we would be confronted with some different screens here. We would have to specify if the data in the file was delimited (each field in our file is separated by commas or tabs) or fixed width (each field in our file starts and ends based on specific character positions). If the file is delimited, we would be asked to specify which delimiter is being used, and can preview if it is correct. If the file is fixed width, we would be asked to insert lines that indicate where one field ends and another begins).

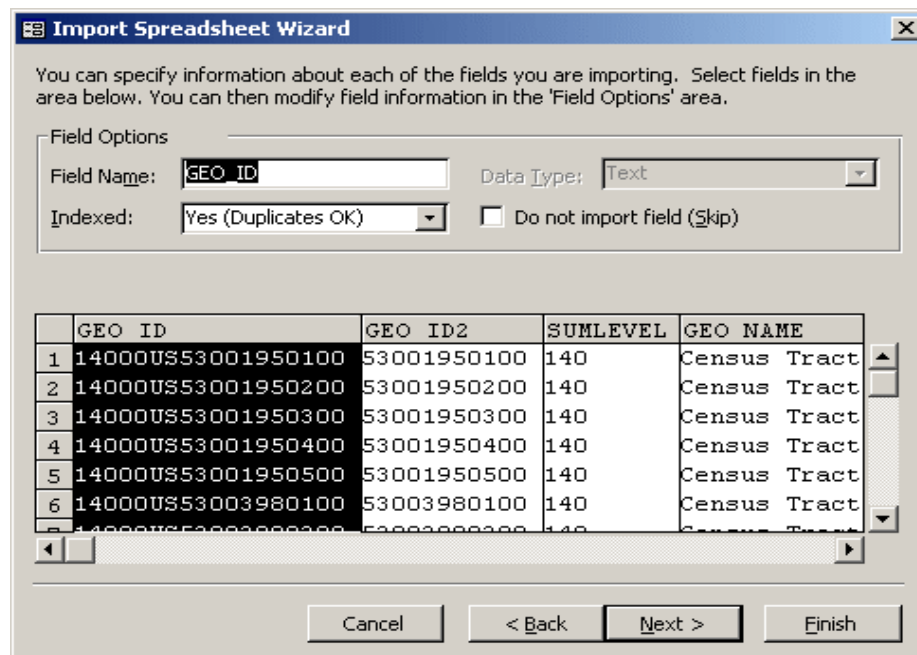
Next, we specify that we want to store the data in a new table (Figure 9).

Figure 9



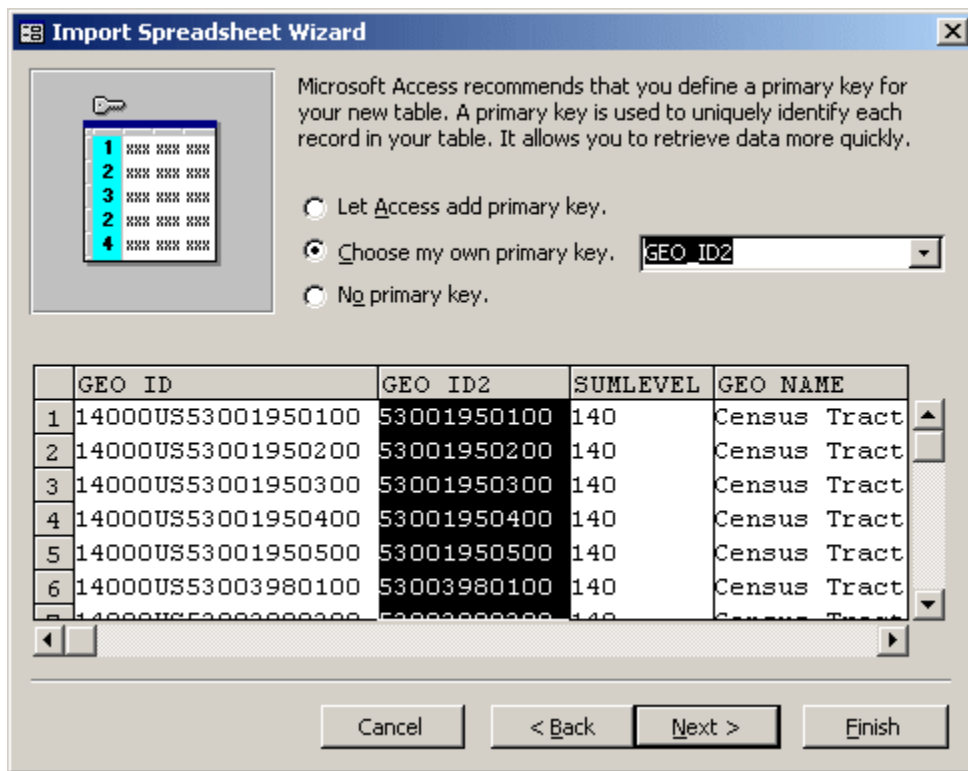
Then we are asked if we want to change any of the field names, data types for the fields, or if we want to exclude particular fields (Figure 10). The simplest thing to do here is to accept the defaults and click *Next*, as we can always make changes to these options later, once the data is imported.

Figure 10



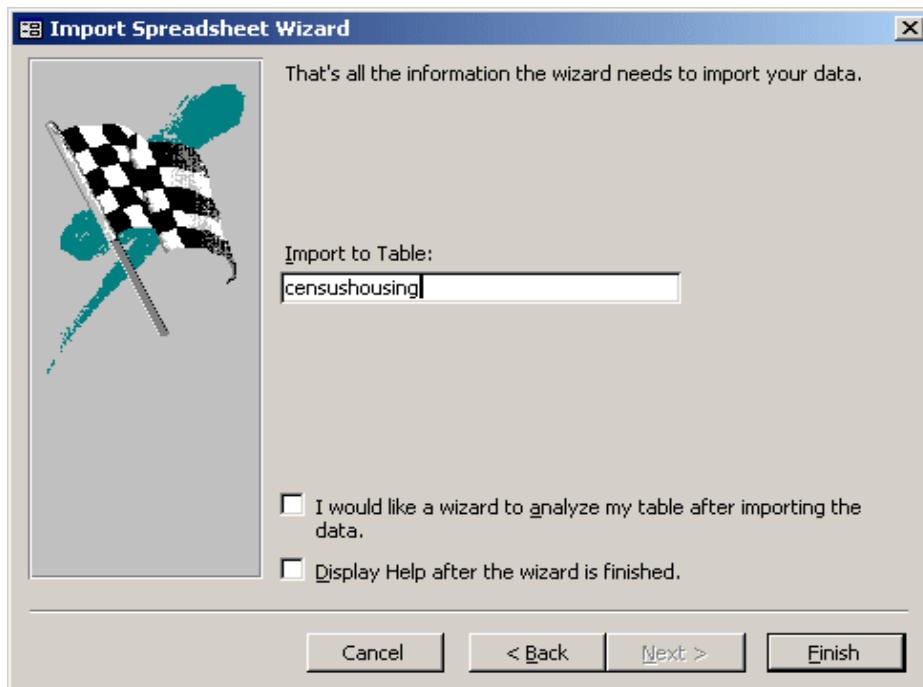
We are then prompted to specify a primary key, which is a number or code that uniquely identifies each individual record (Figure 11). In this case, we know that GEO\_ID2 is the unique number, so we designate it as the primary key. If we did not know or were not sure, we could choose not to assign a key, and could set this option later, once the data is imported.

Figure 11



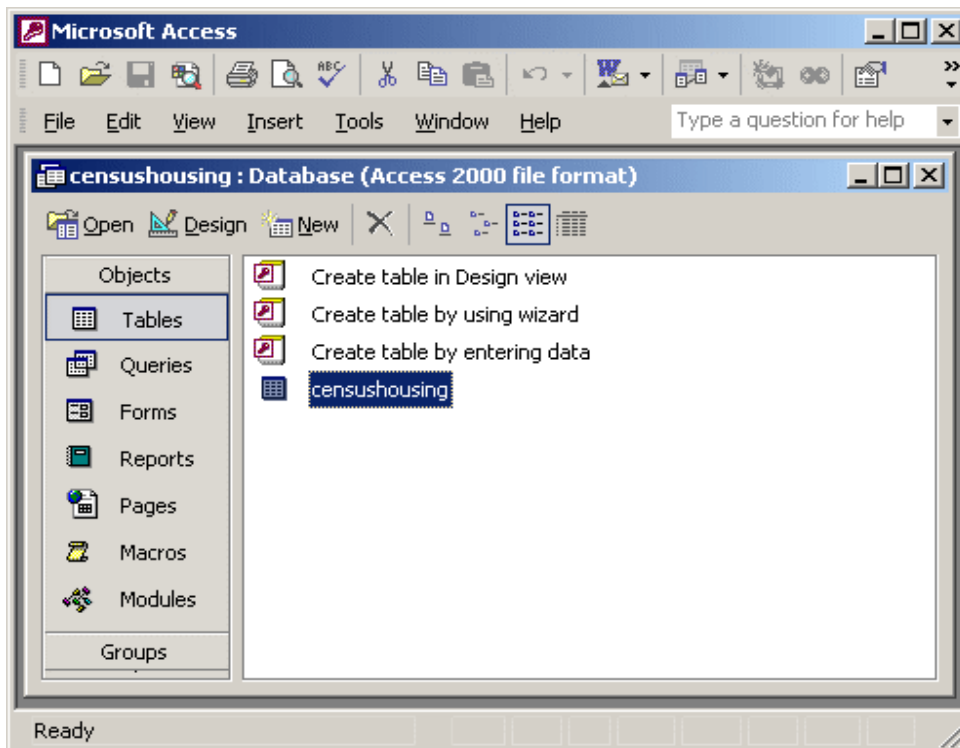
The final screen asks us to name our new table (Figure 12).

Figure 12



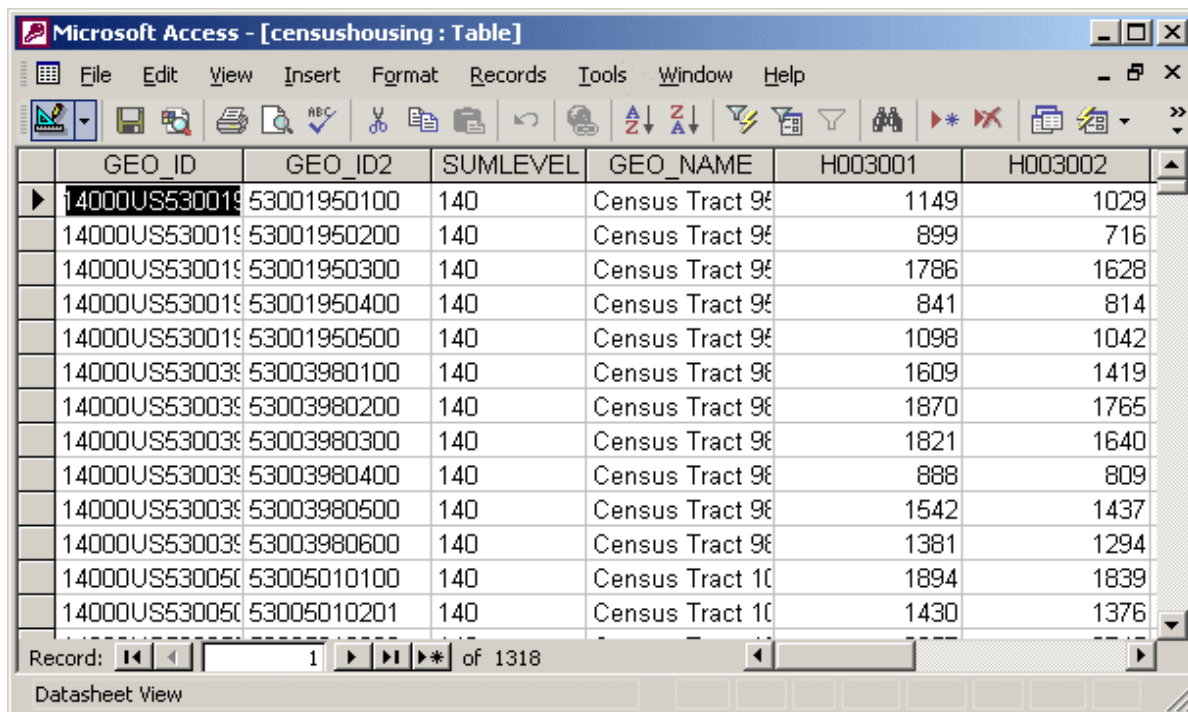
Access then imports our data, which may take awhile depending on how many records there are. When the process is complete, we end up back at the table objects window, and now we have one object in this window – our new *censushousing* table (see Figure 13).

Figure 13



Double-clicking on the table will open it (Figure 14):

Figure 14

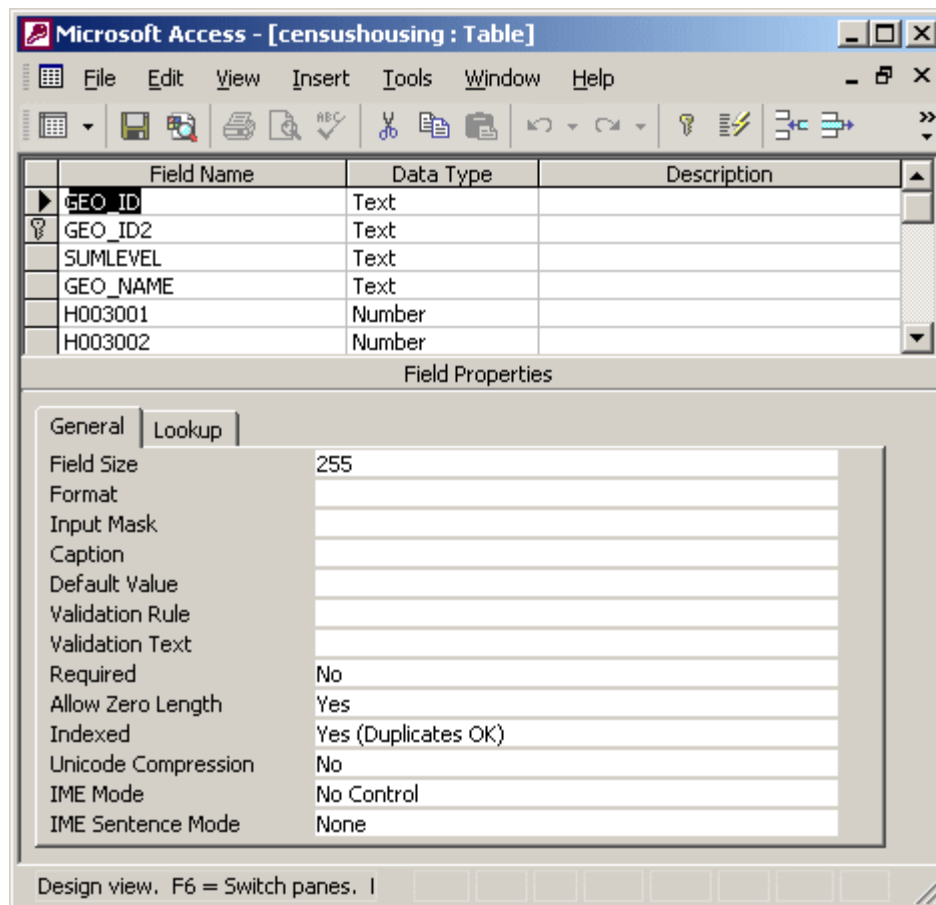




This table view is known as the *Datasheet View*. In this view, we can see our data and the total number of records (at the bottom of the screen). We can also perform many of the functions that we can perform in a spreadsheet, like sorting and filtering records and changing column widths and order. Unlike Excel, Access does not care about column or sort order; all of the data in each record will be kept together regardless of how we move the columns or records. Be careful when altering or deleting records though – once you do it, it cannot be undone!

The other table view is called the *Design View*. You can toggle between the *Datasheet* and *Design* views by with the toggle button on the tool bar in the upper-left hand corner. In the *Datasheet View*, the button has a triangle, ruler and pencil on it. Once you are in the *Design View* the button changes to look like a data table, which will take you back to the *Data View*. You can consider the *Design View* as the behind-the-scenes view for setting the properties for your tables (see Figure 15). You can use this view to change column sort orders and names, but you can also assign primary keys, display settings, and change the data types for each column (text strings, dates, a variety of numbers, etc.)

Figure 15



You may have to alter the data type settings if a column was assigned an incorrect data type. A common example would be an ID field saved as a number instead of text. If you are going to join this table to a shapefile based on a common ID (or to another data table), the joining fields in each table must share the same data type, otherwise the join will fail (for instance, you cannot join string fields to number fields). Another example would be a number that was saved as a string. If you want to perform calculations on a field, it must be saved as a number. When choosing the type of number field, make sure you choose one that can accommodate the proper length. For example, we would not want to save any of our census data as integers, since the range of integers is approximately -32,000 to 32,000, which means we would lose any

data that was greater than 32,000. A long integer would be preferable. If you have numbers that are decimals, make sure that they are not saved as integers; the single or double settings would be necessary to preserve digits to the right of the decimal place. Be careful about changing the data types, as you risk losing data in the conversion process.

If your ultimate goal is to use your table in ArcGIS, you will be able to add it to ArcMap using the same methods you would use for any other type of file. Click the *Add Data* button, and navigate to the folder where the database is stored. Microsoft Access mdb files are represented in ArcGIS file menus as grey cylinders. Once you click on the database, you will be prompted to select the data table within the database that you wish to add to your map.